LA-UR-18-22622
Approved for public release; distribution is unlimited.

Title: Electric Power Outage Forecasting

Author(s): Pasqualini, Donatella
Kaufeld, Kimberly Ann
Dorn, Mary Frances
Vander Wiel, Scott Alan
Backhaus, Scott N.

Intended for: Report

Issued: 2018-03-27

Electric Power Outage Forecasting

Donatella Pasqualini[1], **Kimberly Kaufeld**[1], **Scott Vander Wiel**[1], **and Scott Backhaus**[1]

[1]*Los Alamos National Laboratory*

Correspondence*:
PO Box 1663, MS P939, Los Alamos, NM USA 87545
dmp@lanl.gov +1 505 667 0701

1 INTRODUCTION

Electrical distribution networks are the final leg of the network that moves electric power from generating stations to end users. The vast majority of the circuits in these networks are above-ground and exposed to direct damage from strong winds, broken tree limbs, toppled trees, and flying debris. These systems are also susceptible to ice storms. The combined stress of the weight of ice (icing), the increased wind resistance of the conductors, and broken tree limbs can damage lines, poles, and support structures.

These networks are generally operated in a radial configuration, which makes them susceptible to single-point failures. Some redundancy is built into these networks through circuit switching and alternative power supply points in the network. This redundancy can be overwhelmed by hurricanes and major ice storms that cause widespread damage to these networks. The exposure and limited redundancy of these networks makes them the primary cause of long-term electrical outages following hurricanes and major ice storms.

This paper focuses on electric power outage forecasting for hurricane-force winds and icing conditions in a data-poor environment when distribution network models are not generally available. The remainder of this paper is organized as follows: Section 1 provides additional context for the problem setting and the importance of this capability development to the National Infrastructure Simulation and Analysis Center (NISAC). Section 2 discusses the requirements and limitations relative to the availability of input data, computational speed, and presentation of the results. The literature review in Section 3 discusses existing methods and the utility of these methods to the current problem setting. In Section 4 we describe and mathematically formulate our recommended approach to adapting current methods to NISAC requirements.

1.1 Problem Statement

Distribution utilities use first principles-based power flow solvers on detailed models of their electrical distribution network to simulate effects of local faults or other upsets to their systems. The origin of these faults is generally not specified, rather, the study goal is to design the network redundancy and operations to mitigate these assumed faults. These detailed models could be extended to evaluate or predict outages resulting from heavy, widespread damage scenarios, but several conditions would have to be met. Among these conditions are the availability of the model, geolocation of the components in the model, and sufficient power system component metadata to enable accurate fragility estimations for hurricane-force winds and icing conditions.

Distribution network model availability is a significant challenge. As of 2017, detailed distribution network models and data are not routinely reported to any federal agency. Access to models and data is feasible only via non-disclosure agreements (NDAs) with each distribution utility; there are several thousand distribution utilities in the continental United States (CONUS). Although collecting a few detailed models is feasible and desired for focused studies, this approach is impractical for CONUS-scale studies or even within hurricane-prone areas because of the number of NDAs required.

Even if collecting a large number of detailed distribution models was feasible, the format of these models can be quite diverse; as such, consistent conversion from one format to another is difficult. Many distribution network databases are now fully geolocated, however, in our recent experience, component metadata are sparse or missing entirely and fragility estimates would require many assumptions about component quality and type.

1.2 Importance

The ability to predict electrical outages resulting from hurricane-force winds and ice damage to electrical distribution networks is key for electrical power systems analysis for these extreme events. It is also a key step in the analysis of cascading failures in critical or lifeline infrastructure networks that depend directly or indirectly on electrical power.

2 BOUNDING REQUIREMENTS AND LIMITATIONS

The electrical power forecasting model discussed in this report is intended for use within NISAC. Analysis scope, metrics utilized, available input data, and desired output results are subject to key bounding requirements and limitations, which we use to frame the remainder of the discussion and provide context to the literature review in Section 3.

2.1 Analysis Scope

The electrical power outage forecasting model, associated literature review, and subsequent development of methods considers all CONUS locations. The model's primary function is to estimate the expected electric power outages from hurricane-force winds or icing conditions at a county-level spatial scale. NISAC lacks access to detail distribution network models so these estimates will be made using area-based statistical models that are trained on utility-reported historical outage data. Although not explicitly incorporated into the model, this class of methods will capture the effects of distribution network switching and redundancy and network hardening if these training data captures these effects.

2.2 Input Data: Scope and Limitations

The scope of the electrical power outage forecasting model is CONUS at the county level, therefore, both static and dynamic input data to the forecasting tool should be uniformly and publicly available for all of CONUS at that scale. The input data must be routinely maintained and updated by the original data provider or NISAC so that the model can be run as dynamic conditions change; e.g., storm track and precipitation; and static or slowly changing input data evolve, e.g., population, soil moisture, and vegetation.

2.2.1 Model Training Data

The forecasted response variable is the number of customers without electric power in each county. To train the model, we intend to use historical outage data from the EAGLE-I database (EAGLE-I), which collects near real-time electric power outage data. The database includes outages reported by the utilities for most of the counties in CONUS. The predictor variables will be quite diverse and include hurricane gust wind speed, population, land cover type, standardized precipitation index, and soil moisture index.

2.2.2 Model Implementation Data

To implement the model for real-time events, we require the dynamic set of predictor variables to be available and updated in real time.

2.3 Analysis Outputs and Presentation

The output of the predictive model will be the number of customers without electric power within each county in CONUS. These output data will be available within LANL's AWS computing environment via the AGAVE presentation map layer or its successors. These data will also be available in table form.

2.4 Computational Requirements and Cost

We will train the model in an offline setting where there are no strict requirements on computational speed. For operational use of the model, the model should be able to compute predictions in less than 5 minutes because there are several other computational processes to complete in the 15-minute window allowed for model output update following the update of input data by other providers.

3 LITERATURE REVIEW

Considering the key bounding requirements and limitations discussed in Section 2, we review the existing literature and discuss its applicability. Our literature review revealed specific research focused on forecasting electric power outages due to wind. The review also identified key outage forecasting methods and parameters. LANL did not find significant literature for outage forecasting due to icing conditions.

3.1 Generalized Linear Models

Linear models assume that the relationship between the response, y, and predictors, x_j, is linear and that the form of the model is $y_i = \sum_j^k \beta_j x_{ij} + \epsilon_i$ where $\epsilon \overset{iid}{\sim} N(0, \sigma^2)$. Outage data consist of counts of the number of customers without electric power resulting from a hurricane, wind, or ice event. Count data cannot be negative and the dependence of count data on the inputs is not expected to vary in linear fashion over a wide range of the predictors, e.g., the number of customers without power should saturate at very high wind speeds. Normal linear model theory is not expected to perform well in this setting.

An alternative that overcomes some of these issues is a generalized linear model (GLM). GLMs allow for more complex dependence of the predicted (dependent) variable Y on the random (independent) variables x_i by introducing an intermediate linear predictor η and a link function g. The relationship between the linear predictor variable and the predictor variables x_i is

$$\eta = \beta_0 + \sum_i \beta_i x_i,$$

where the β_i are the model parameters fitted to the data. The dependent predicted variable Y is related to the intermediate linear predictor using the link function by $E[Y] = g^{-1}(\eta)$ where E indicates the expected value of Y. Different canonical link functions g are used in conjunction with different distributional assumptions, including Gaussian, Poisson, and Gamma. In the case of the Poisson and negative binomial models, the expected value of these distribution is $E[Y] = \exp(\eta)$ corresponding to a link function $g(\eta) = \log(\eta)$. That is $E[Y] = \exp(\beta_0 + \sum_i \beta_i x_i)$.

GLMs were some of the first models used to consider electric power outages in Guikema et al. (6); Liu et al. (18); Guikema (5); Han et al. (10, 11). Based upon the literature, a comparison of Poisson and negative binomial models suggests that the negative binomial distribution is the most appropriate because the data are over-dispersed, i.e., the mean is not equal to the variance as assumed with a Poisson distribution. The negative binomial model used to estimate the geographical distribution of power outages in North and South Carolina showed the best predictor variables were the number of transformers in the area, company affected, maximum gust wind speed, and hurricane effect.

3.2 Generalized Additive Models

Generalized additive models (GAMs) are similar to GLMs, but they relax the assumption of linearity in the intermediate predictor, η. Rather than using individual predictor variables or linear combinations of predictors to form the linear predictor, the linear predictor is related to the predictors through the use of smoothing splines f_i

$$\eta = \beta_0 + \sum_i f_i(x_i),$$

e.g., cubic regression splines or thin plate splines (12). Although flexible, GAMs can lead to over fitting of the data and can be more difficult to fit because of the selection of the smoothing functions f_i.

Guikema (5) provided an overview of GAMs and GLMs for predicting electric power outages. In particular, Han et al. (10) used GLMs and GAMs to estimate power outages resulting from hurricanes and found that the predictive performance was better for the GAM relative to a GLM, based upon holdout validation. Guikema et al. (9) compared GLMs and GAMs to data mining methods, and found that GAMs and data mining methods had better predictive accuracy then GLMs, however, the aggregation of the data can make it difficult to create detailed geographic damage estimates for all types of models.

3.3 Classification and Regression Trees

Classification and regression trees (CART) are the foundation of several ensemble-based methods for constructing predictive models from data. These methods do not make any assumptions about the underlying functional distribution relating the response and the predictor variables. Rather, the models are constructed by iteratively partitioning the predictor data space into sub-regions and fitting a predictive model within each sub-region. The partitioning results in a tree structure are called a decision tree.

In classification trees, the predicted value is not continuous; it is a type or a label, e.g., red or blue. The classification tree predicts the most prevalent type in each sub-region of the predictor data space. In contrast, a regression tree predicts the mean of the value of the dependent variable in each sub-region of the predictor data space. Regression trees are more applicable to predicting electric power outages, however, if the training data are limited, limited, single regression trees have a tendency to overfit.

A random forest is a non-parametric data mining ensemble approach developed by Breiman (1) that mitigates the tendency of regression trees to overfit. It creates many regression trees using random subset samples from the full training data set. Each regression tree is used to make an independent prediction of the dependent variable; these predictions are averaged to make the ensemble prediction. By randomly sampling from the training data, the regression trees are approximately uncorrelated and unbiased, which results in better aggregate performance.

There are several extensions and generalizations of the random forest approach, including boosted gradient tree regression, Bayesian additive regression trees (BART), and quantile regression forests (QRF). Boosted gradient tree regression (4) is an additive decision tree that builds a series of shallow regression trees where the predictors for each tree are residuals from the previous tree. As more trees are produced, the residuals of the previous small trees can become a good predictive model. These type of trees are considered “weak learners” because there is not as much information in each of the trees, although the average over all trees can be a powerful predictor.

BART is another additive method based on shallow trees of limited complexity that are weak learners. The model fit and model prediction is done by a Markov chain Monte Carlo algorithm.

QRF, developed by Meinshausen (19), is a non-parametric method to predict conditional quantiles of the dependent variable using an approach based on a random forest model. Compared to a traditional random forest that predicts the mean, QRF provides more information about the distribution of the dependent response variable as a function of the independent predictors.

Guikema et al. (7) and Nateghi et al. (24, 22) used the random forest method to predict electric power outages at different spatial resolutions for multiple hurricanes. The model captured the nonlinear structure of the outage data and produced reasonably accurate prediction estimates for both power outage durations and number of outages. He et al. (13) used quantile regression models and compared them to other machine learning methods to predict electric power outages at a 2-kilometer scale using simulated weather data. They also compare the model to random forests by Tonn et al. (26), finding that both models had good out-of-sample accuracy, but that quantile regression was able to provide additional information on the uncertainty and accuracy of the estimates. Nateghi et al. (23) used BART to predict power outage durations after hurricanes, while He et al. (13) used BART to predict the number of outages due to a hurricane with good predictive accuracy. An additional comparison by Nateghi et al. (24) between BART and random forests found the random forest model outperformed BART, producing better predictive accuracy for power outage durations.

3.4 Multivariate Adaptive Regression Splines

Friedman (3) developed multivariate adaptive regression splines (MARS). Compared to GLM or GAM, MARS is a flexible regression method that does not make any assumptions about the underlying functional form relating the response and the predictor variables. The MARS method constructs a basis consisting of hinge functions and their products in an adaptive way by automatically selecting appropriate variables and knots for the set of functions. The result is a continuous hinged surface across potentially many predictor variables. The predictions for a set of predictors x_i are determined from a given configuration of knots using the training data.

Nateghi et al. (23) compared ensemble methods, BART and CART, and MARS to determine which model had the best predictive performance for power outage duration. Of these methods, BART, followed by MARS, had the best prediction accuracy and predicted power outage duration reasonably close to the actual power duration reported.

3.5 Bayesian Networks

Mensah and Dueñas-Osorio (21) used Bayesian networks to predict electric power outage due to hurricane winds in Harris County, Texas. Bayesian networks are probabilistic graphical models that represent a set of predictors and their dependencies by a directed graph, rather than a tree. The model framework depends on the component fragilities and topology of the electric power grid. The model made reasonably accurate predictions of outages at the ZIP code level, however, it is based upon the distribution network structure and exceeds the limitations noted in Section 2.2.

3.6 Summary

Table 1 summarizes the different statistical methods, predictors, and data sources found in the literature review. GLMs using the Poisson distribution are generally a good choice, however, if the data are over-dispersed, the negative binomial distribution is more appropriate because it accounts for the variance being greater than the mean. GAMs are more flexible due to their nonparametric nature, but are susceptible to overfitting and can be more difficult to fit due to the need to find a smoothing function.

Regression trees are relatively easy to use, especially when there are several predictors. They account for interactions of categorical predictors, i.e., tree species, in a more flexible way, the relationship between the predictors and response do not need a linear relationship. Ensemble regression tree methods are superior to single-tree methods and include BART, random forest, and quantile regression forests. Guikema et al. (7) and Nateghi et al. (24, 22) found that predicting electric power outages based upon previous storms was better using these QRF and random forest when compared to GLMs and GAMs. MARS provides reasonable estimates and prediction accuracy with slightly less accuracy compared to BART, but has been implemented only in assessing power outage durations.

Bayesian networks provide a useful machine learning tool to predict relationships between nodes; however, they can be computationally expensive and have only been investigated for settings where electrical distribution network data were available.

Table 1. Summary statistical methods for forecasting electric power outages resulting from hurricanes.

Model	Predictors
GLM	Maximum 3-second gust wind speed, wind duration, soil moisture, precipitation, land cover database, tree type, company affected, 7-day rainfall (in), soil type **Sources:** National Climate Data Center (NCDC), Forest Land Distribution Data, STATSGO, National Land Cover Database, hurricane wind model **Citations:**Li et al. (14); Liu et al. (16, 18); Han et al. (10) Liu et al. (17, 18); Guikema and Quiring (8)
GAM	maximum three-second gust wind speed, wind duration, soil moisture, precipitation, land cover database **Sources:** Duke Energy distribution, Census (population density), GIS-based hurricane wind field simulation model, NCDC, Landsat, Forest Land Distribution Data, U.S. Department of Agriculture Natural Resources Conservation Service State Soil Geographic Database **Citations:** Guikema and Quiring (8); Han et al. (11); Guikema et al. (9)
Tree	Tree and wind gusts, soil moisture, precipitation, land use, tree trimming, soil moisture, temperature, total accumulated precipitation, and leaves on trees. **Sources:** Census (population density), GIS-based hurricane wind field simulation model, NCDC, Landsat, Forest Land Distribution Data, STATSGO, Standardized Precipitation Index, tree trimming, power grid data, WRF model **Citations:** He et al. (13); Wanik et al. (27); Nateghi et al. (22, 24), Guikema et al. (7); Nateghi et al. (23); Guikema et al. (9)
MARS	Maximum 3-second gust wind speed, wind duration, soil moisture, precipitation, land cover database, tree trimming, soil moisture, precipitation, land use. **Sources:** Census (population density), GIS-based hurricane wind field simulation model, NCDC, Landsat, Forest Land Distribution Data, U.S. Department of Agriculture Natural Resources Conservation Service State Soil Geographic Database, power grid data **Citations:** Nateghi et al. (23)
Bayesian	Private power grid data with transmission substations and transmission lines, tree damage **Citations:** Mensah and Dueñas-Osorio (21)

4 PROPOSED METHOD

This literature review has revealed that, given the limitations on input data available to NISAC, ensemble regression tree methods such as random forest, BART, and QRF are preferable to Bayesian network approaches. When compared to GLM or GAM, ensemble regression trees are better adapted to handle a diverse set of predictor variables, including variables with discrete or integer representation, e.g., tree species. Several studies have also shown that ensemble regression tree methods have better prediction accuracy than GLMs, GAMS, or MARS. Although training ensemble regression tree methods may be computationally slow, these methods are computationally efficient in implementation.

Based on these comparisons, we recommend ensemble regression tree methods for the next generation of NISAC's electric power outage prediction capability. Among the different choices for ensemble regression tree methods, we suggest development of random forest over QRF and BART because of its relative maturity in the outage forecasting literature and its simplicity.

4.1 Random Forest Model

Random forest models are estimated using the "randomForest" package in R by Liaw et al. (15) and QRF are estimated using the "quantregForest" package by (20) in R, (25). Here we provide a qualitative and semi-quantitative discussion that summarizes the technical details of these models.

Ensemble regression tree (or forest) models use an underlying general modeling framework of the form

$$Y = g(x) + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \tag{1}$$

where inferences are made for the response Y using an unknown function g, which is approximated as $g(x) = E(Y|x)$, using a p dimensional vector of inputs $x = (x_1, \cdots, x_p)$. We model $g(x)$ as a linear combination of regression trees. There are different methods for combining a set of regression tree models, such as boosting, bagging, random forests, and quantile regression forests. Each uses slightly different linear combination of trees. We focus on random forests, but also include a discussion of QRFs because the mathematics are quite similar.

Qualitatively, in a random forest of regression trees, $g(x)$ is approximated from a sum of m regression trees

$$g(x) = (1/m)\sum_{i=1}^{m} f_i(x), \tag{2}$$

where $f_i(x)$ is a single regression tree.

For each tree, $f_i(x)$, and each node in the tree, we use randomization to select which variable x_i to partition the data space x. Each tree $f_i(x)$ is trained on a different random subset of the training data. Further, the regressor variable associated with each node of each tree is selected at random. This recursive random partitioning selects a random subset of predictor variables to avoid correlations between the tree structures. The prediction for a given tree at a point x is a weighted average of the original observations with weights depending on x. The prediction of the random forest is a straight average over all the trees (Eq. 2).

Breiman (1) provides a more rigorous discussion, which we summarize here. Each tree in the random forest is denoted by $T(\theta_j)$ where θ_j is a vector of parameters associated with the variables x_i used to partition the data at each node. In each tree $T(\theta_j)$, there is a set of leaf nodes denoted $\ell(x, \theta_j)$ that covers a rectangular sub-region of the predictor space $X = x$ determined by the parameters θ_j. For any predictor in this leaf's sub-region, the prediction of the tree $T(\theta_j)$ is an average over all the observed values in the training data set that also fall in leaf $\ell(x, \theta_j)$, i.e.,

$$\hat{\mu_j}(x) = \sum_{i}^{n} w_i(x, \theta_j) Y_i \tag{3}$$

where $w_i(x, \theta_j)$ is a weight vector, summing to one, and equal to a positive value if an observation X_i is part of a leaf or 0 if it is not.

The prediction of the random forest is an approximation of the conditional mean, $E(Y|x)$, constructed by averaging over the predictions of m trees

$$E(Y|X = x) = \hat{\mu} = \frac{1}{m}\sum_{j=1}^{m} \hat{\mu_j}(x).$$

Quantile regression forests developed by Meinshausen (19) are a generalization of random forests. Instead of classifying only the mean, they estimate the conditional quantiles based upon an underlying random forest model. The weighted observations can also provide a good approximation of the full conditional distribution. The conditional distribution function of Y given $X = x$ is

$$\begin{aligned} F(y|X = x) &= P(Y \le y|X = x) \\ &= E(1_{\{Y \le y\}}|X = x) \end{aligned}$$

Using the weights, $w_i(x)$, from random forests, we use QRF to approximate the weighted mean over the observations of $1_{\{Y \le y\}}$ as,

$$\hat{F}(y|X = x) = \sum_{i=1}^{n} w_i(x) 1_{\{Y \le y\}}.$$

Instead of keeping the mean of the observation, as in a random forest, QRFs keep the values of all observations to assess the conditional distribution.

4.2 Training and Characterization

We will train the random forest model using historical, county-level electric power customer outage data for hurricane events taken from the EAGLE-I database. We will use the following process to train and characterize the model:

1. Subdivide the original training data randomly into 10 equal-size subsets and create 10 new training sets by sequentially holding out one of the subsets
2. Train 10 random forest models, one for each of the 10 new training data sets, using standard methods in the software package R
3. Calculate the prediction error for the hold-out data for each of the 10 random forest models
4. Generate the final predictive model by averaging the results from the 10 random forest models

The prediction error for each of the 10 random forest models will be computed using two metrics–Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The MAE is an absolute measurement of the difference between the predicted value and the true value, which is calculated by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|.$$

The RMSE measures the magnitude of error of the predicted values and variability which is defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

We use the following predictor data, based upon the findings of Nateghi et al. (22), to develop the random forest models: hurricane gusts, wind speeds, seven-day precipitation accumulation, population, land cover type, elevation, standardized precipitation index, and soil moisture index. We will use the MAE and RMSE metrics to compare models that use all the predictors with models that use different subsets of predictors to develop the best subset of predictor variables.

REFERENCES

[1]Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

[EAGLE-I]EAGLE-I. Energy infrastructure mapping system and the real time energy monitoring dashboard. https://eagle-i.doe.gov. 2017-04-18.

[3]Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67.

[4]Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.

[5]Guikema, S. D. (2009). Natural disaster risk analysis for critical infrastructure systems: An approach based on statistical learning theory. *Reliability Engineering & System Safety*, 94(4):855–860.

[6]Guikema, S. D., Davidson, R. A., and Liu, H. (2006). Statistical models of the effects of tree trimming on power system outages. *IEEE Transactions on Power Delivery*, 21(3):1549–1557.

[7]Guikema, S. D., Nateghi, R., Quiring, S. M., Staid, A., Reilly, A. C., and Gao, M. (2014). Predicting hurricane power outages to support storm response planning. *IEEE Access*, 2:1364–1373.

[8]Guikema, S. D. and Quiring, S. M. (2012). Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data. *Reliability Engineering & System Safety*, 99:178–182.

[9]Guikema, S. D., Quiring, S. M., and Han, S.-R. (2010). Prestorm estimation of hurricane damage to electric power distribution systems. *Risk analysis*, 30(12):1744–1752.

[10]Han, S.-R., Guikema, S. D., and Quiring, S. M. (2009a). Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models. *Risk analysis*, 29(10):1443–1453.

[11]Han, S.-R., Guikema, S. D., Quiring, S. M., Lee, K.-H., Rosowsky, D., and Davidson, R. A. (2009b). Estimating the spatial distribution of power outages during hurricanes in the gulf coast region. *Reliability Engineering & System Safety*, 94(2):199–210.

[12]Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC press.

[13]He, J., Wanik, D. W., Hartman, B. M., Anagnostou, E. N., Astitha, M., and Frediani, M. E. (2016). Nonparametric tree-based predictive modeling of storm outages on an electric distribution network. *Risk Analysis*.

[14]Li, H., Treinish, L. A., and Hosking, J. R. (2010). A statistical model for risk management of electric outage forecasts. *IBM Journal of Research and Development*, 54(3):8–1.

[15]Liaw, A., Wiener, M., and Liaw, M. A. (2015). Package ‘randomforest’.

[16]Liu, H., Davidson, R. A., and Apanasovich, T. V. (2007). Statistical forecasting of electric power restoration times in hurricanes and ice storms. *IEEE Transactions on Power Systems*, 22(4):2270–2279.

[17]Liu, H., Davidson, R. A., and Apanasovich, T. V. (2008). Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms. *Reliability Engineering & System Safety*, 93(6):897–912.

[18]Liu, H., Davidson, R. A., Rosowsky, D. V., and Stedinger, J. R. (2005). Negative binomial regression of electric power outages in hurricanes. *Journal of infrastructure systems*, 11(4):258–267.

[19]Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7(Jun):983–999.

[20]Meinshausen, N. (2015). *quantregForest: Quantile regression forests.* R package version 0.2-3.

[21]Mensah, A. F. and Dueñas-Osorio, L. (2014). Outage predictions of electric power systems under hurricane winds by bayesian networks. In *Probabilistic Methods Applied to Power Systems (PMAPS), 2014 International Conference on*, pages 1–6. IEEE.

[22]Nateghi, R., Guikema, S., and Quiring, S. M. (2014a). Power outage estimation for tropical cyclones: improved accuracy with simpler models. *Risk analysis*, 34(6):1069–1078.

[23]Nateghi, R., Guikema, S. D., and Quiring, S. M. (2011). Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes. *Risk analysis*, 31(12):1897–1906.

[24]Nateghi, R., Guikema, S. D., and Quiring, S. M. (2014b). Forecasting hurricane-induced power outage durations. *Natural hazards*, 74(3):1795–1811.

[25]R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

[26]Tonn, G. L., Guikema, S. D., Ferreira, C. M., and Quiring, S. M. (2016). Hurricane isaac: a longitudinal analysis of storm characteristics and power outage risk. *Risk analysis*.

[27]Wanik, D., Anagnostou, E., Hartman, B., Frediani, M., and Astitha, M. (2015). Storm outage modeling for an electric distribution network in northeastern usa. *Natural Hazards*, 79(2):1359–1384.